

Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L | week 08

Pierre-Luc Germain

Today's plan

- Debriefing on the assignment
- Follow-up on last week
 - Visualizing signals around selected motifs
- Motif accessibility analysis
 - chromVAR, and working with SummarizedExperiment objects
- Differential motif accessibility analysis

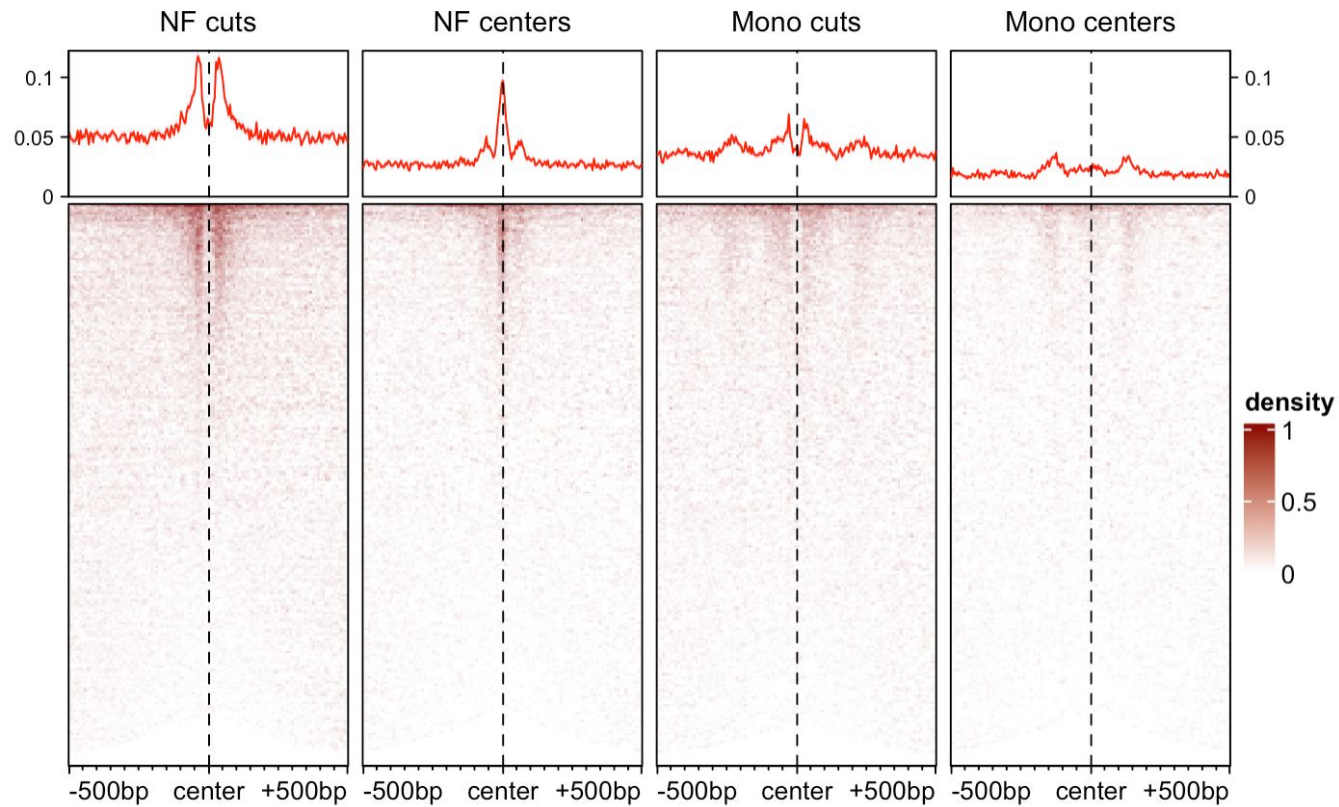
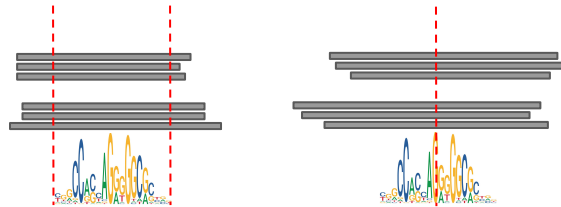
Debriefing on the assignments

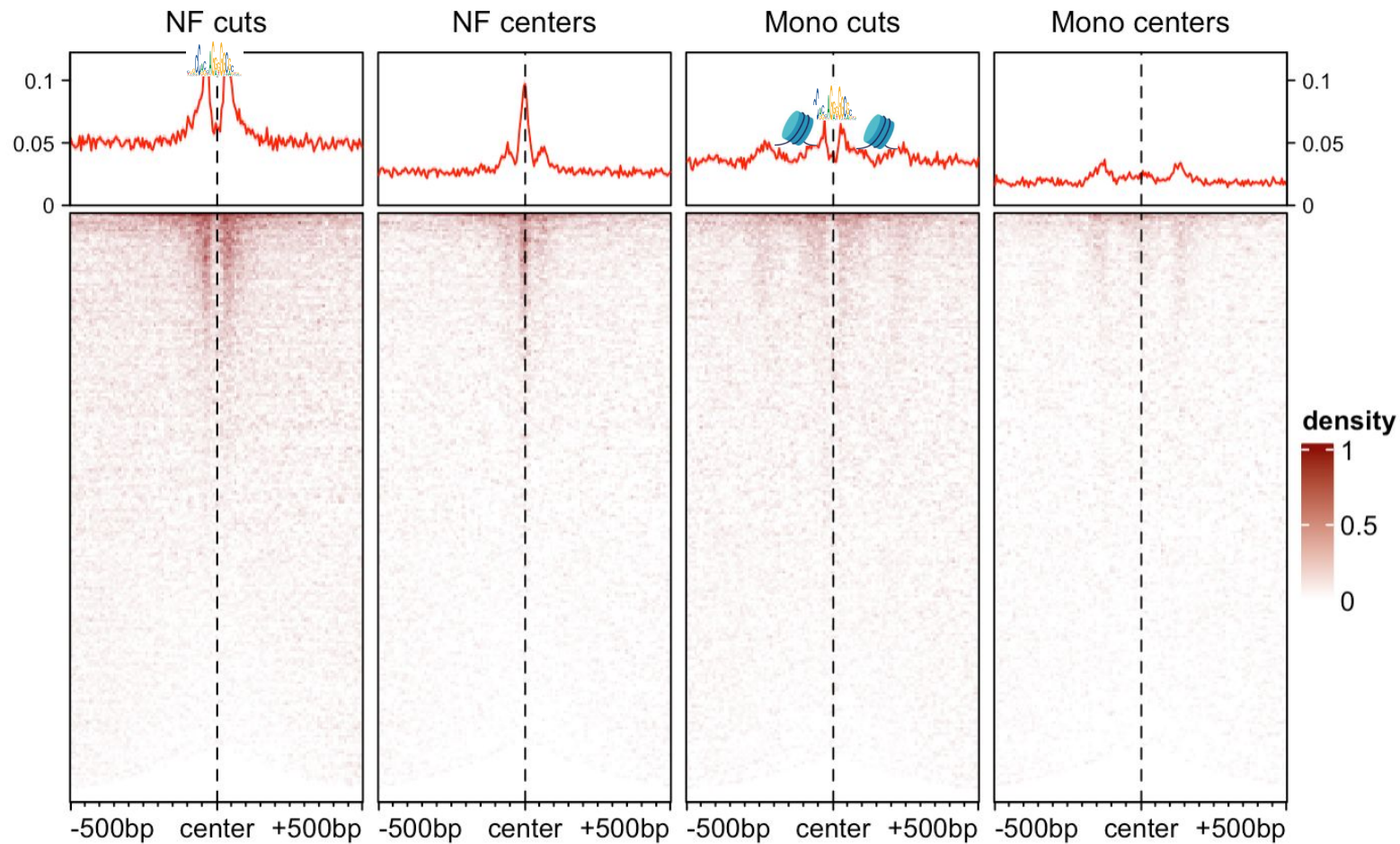
Obtaining **mono**-nucleosome fragments - **cuts**:

```
bam2bw(bam, output_bw = "mono_cuts.bw", paired=TRUE, binWidth=5L, minFragLength=140, maxFragLength=220,  
type="ends", forceSeqlevelsStyle = "Ensembl")
```

Obtaining **mono**-nucleosome fragments - **centers**:

```
bam2bw(bam, output_bw = "mono_cuts.bw", paired=TRUE, binWidth=5L, minFragLength=140, maxFragLength=220,  
type="centers", forceSeqlevelsStyle = "Ensembl")
```





Plotting arguments

```
plotEnrichedHeatmaps(smb, trim=0.95, colors = c("white","darkred"), minRowVal=15)
```

```
Error in if (r[[1]] == r[[2]]) r <- range(unlist(lapply(ml, range))) :
```

```
missing value where TRUE/FALSE needed
```

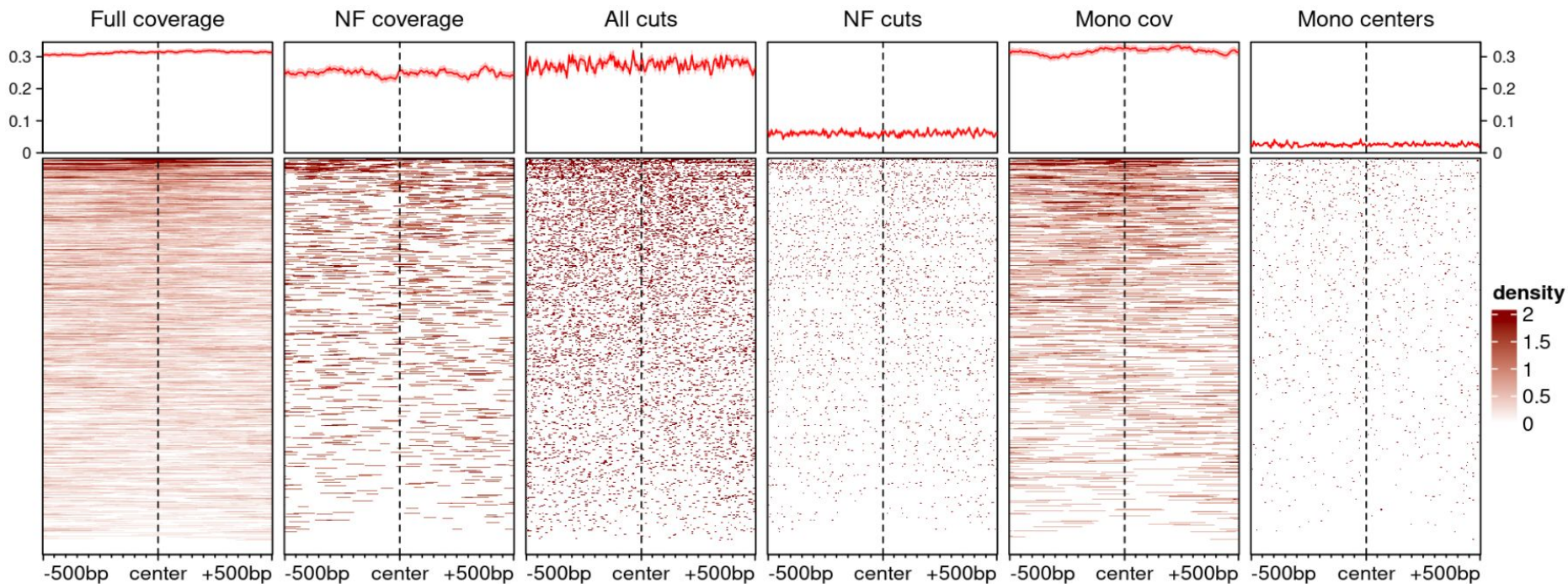
=> After **filtering** for a minimal count of 15 **per row** (i.e. region), there might be no / too little signal across the **motif** for some **TFs** for the respective experiment. This leads to no rows being included in the plot.

```
plotEnrichedHeatmaps(smb, trim=0.95, colors = c("white","darkred"), minRowVal=0)
```

=> Works because there is no filtering (just **very little** signal)

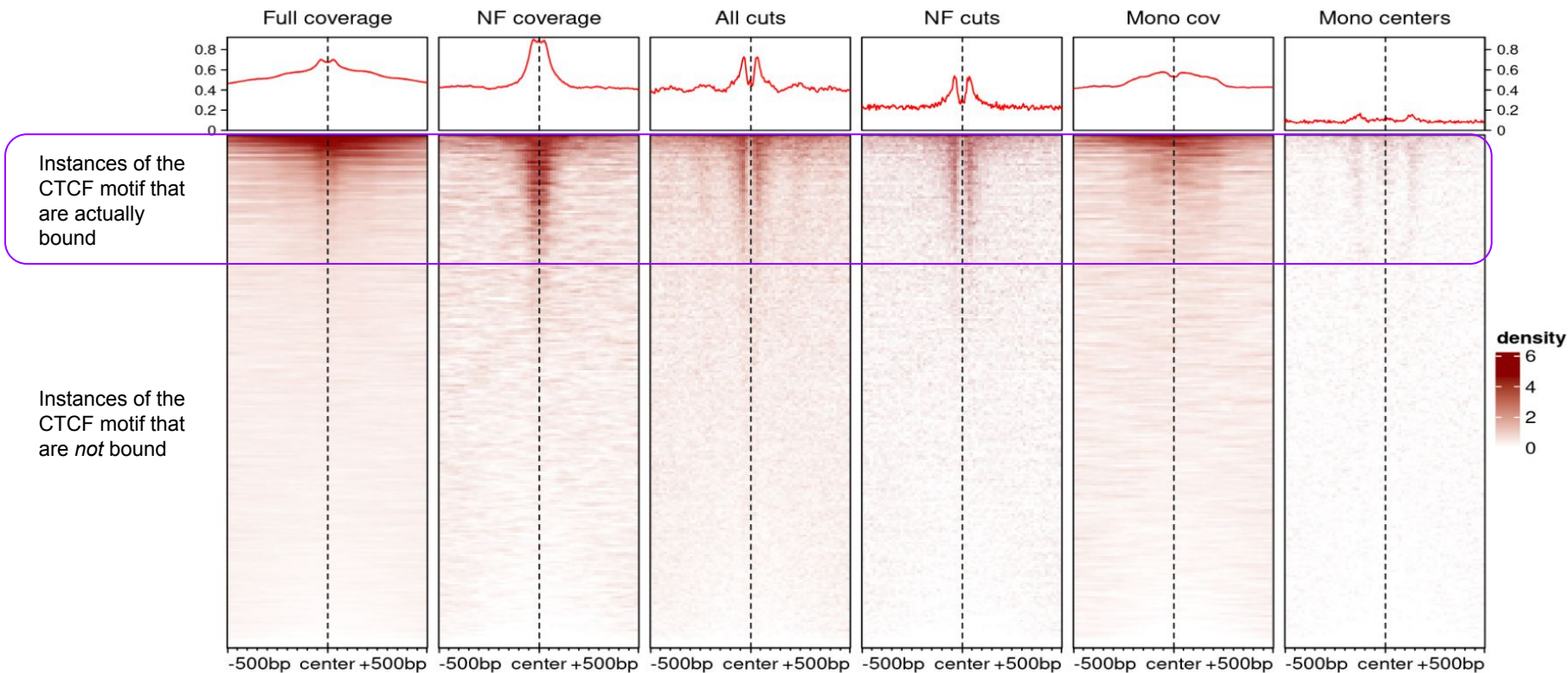
Follow-up on last week

Signal around occurrences of the GATA1 motif in mESC, where the factor is *not* expressed



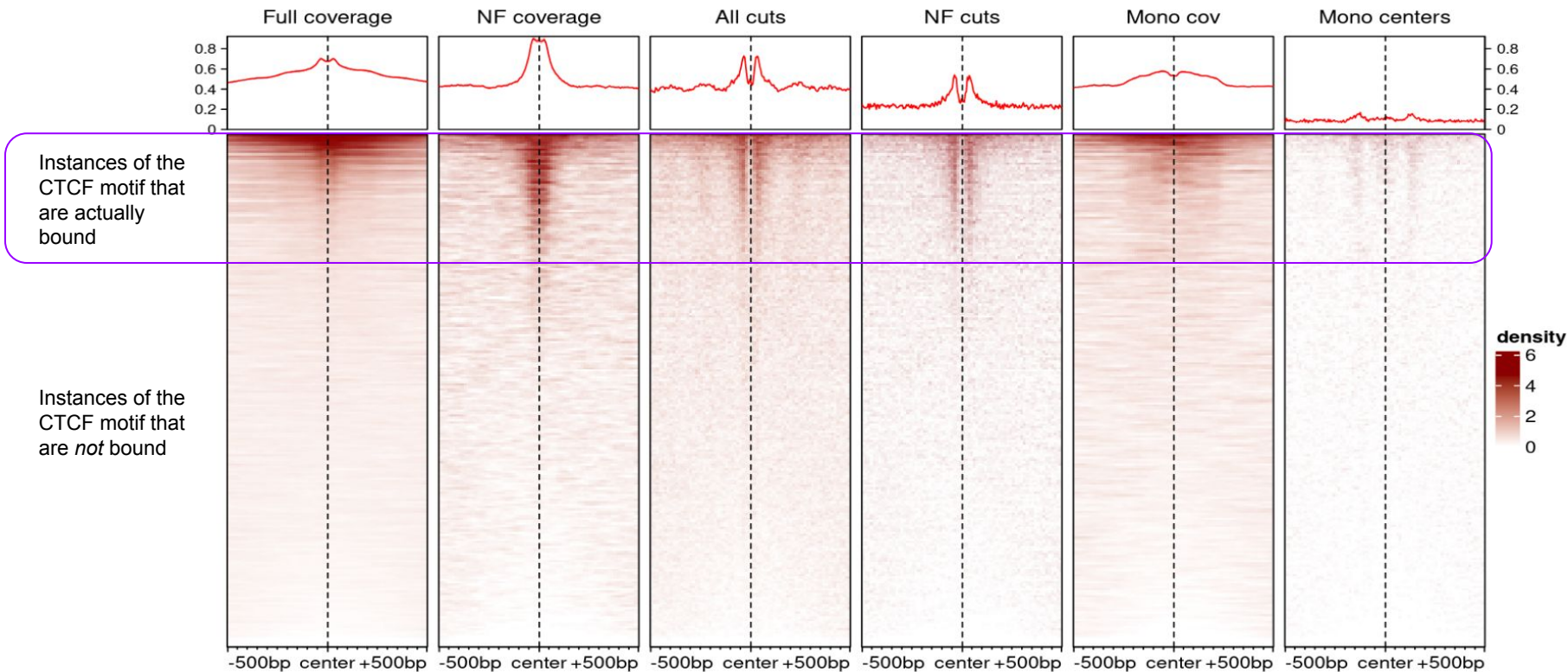
Follow-up on last week

Signal around occurrences of the CTCF motif in mESC



Follow-up on last week

Signal around occurrences of the CTCF motif in mESC



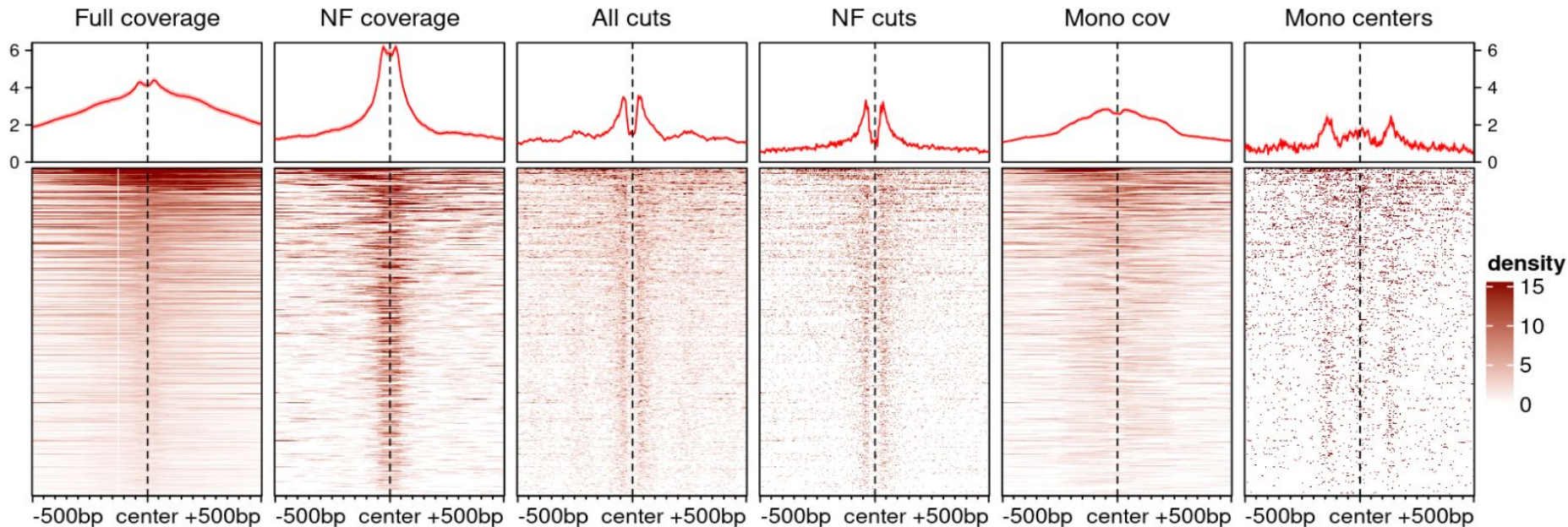
While this is useful to get an overall picture, most of the time we want to concentrate on the actual sites bound

Follow-up on last week

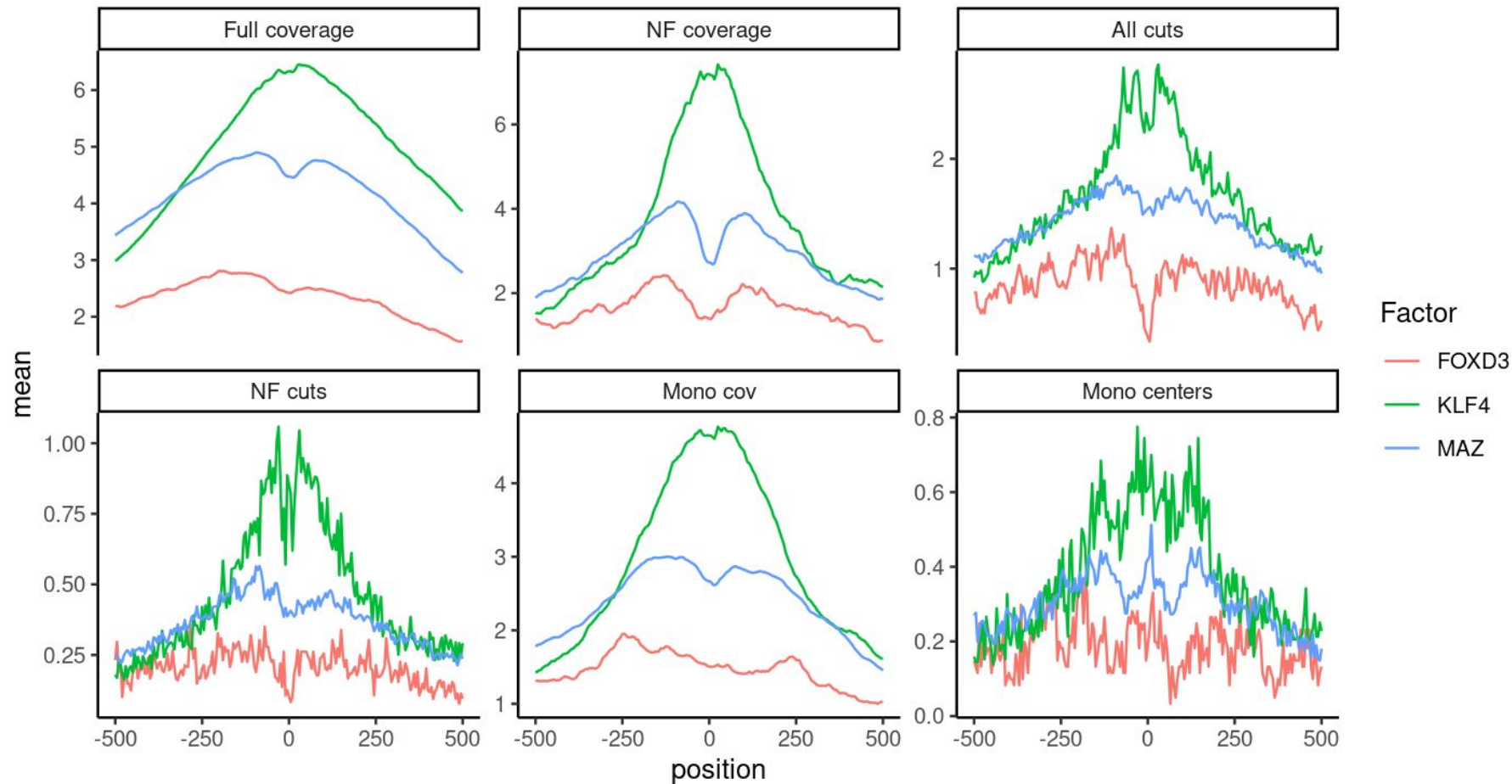
To concentrate on these actual binding events, we can simply restrict the motif occurrences to those within ATAC peaks:

```
moi <- moi[overlapsAny(moi, peaks)]
```

Signal around occurrences of the CTCF motif *in ATAC-seq peaks*



Follow-up on the assignment



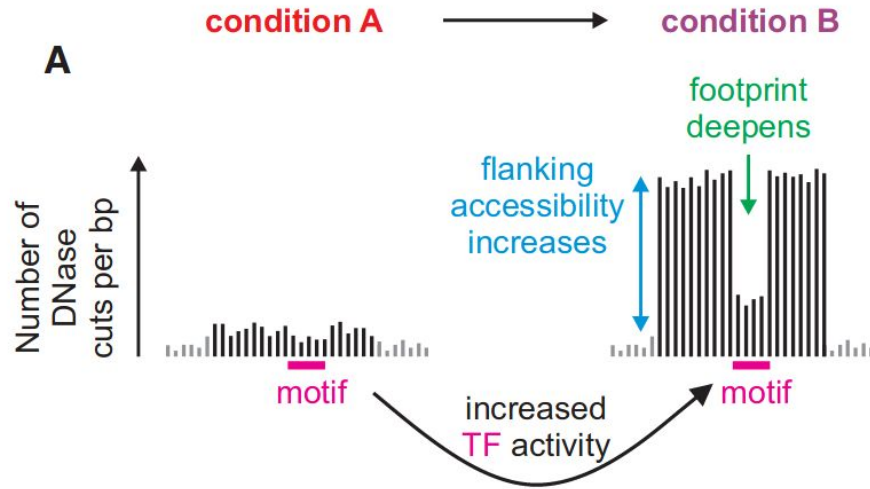
Motif accessibility analysis

What if, rather than looking at a few motifs at a time, we could simply quantify the accessibility/activity of every motif?

Several methods have been developed to do this.

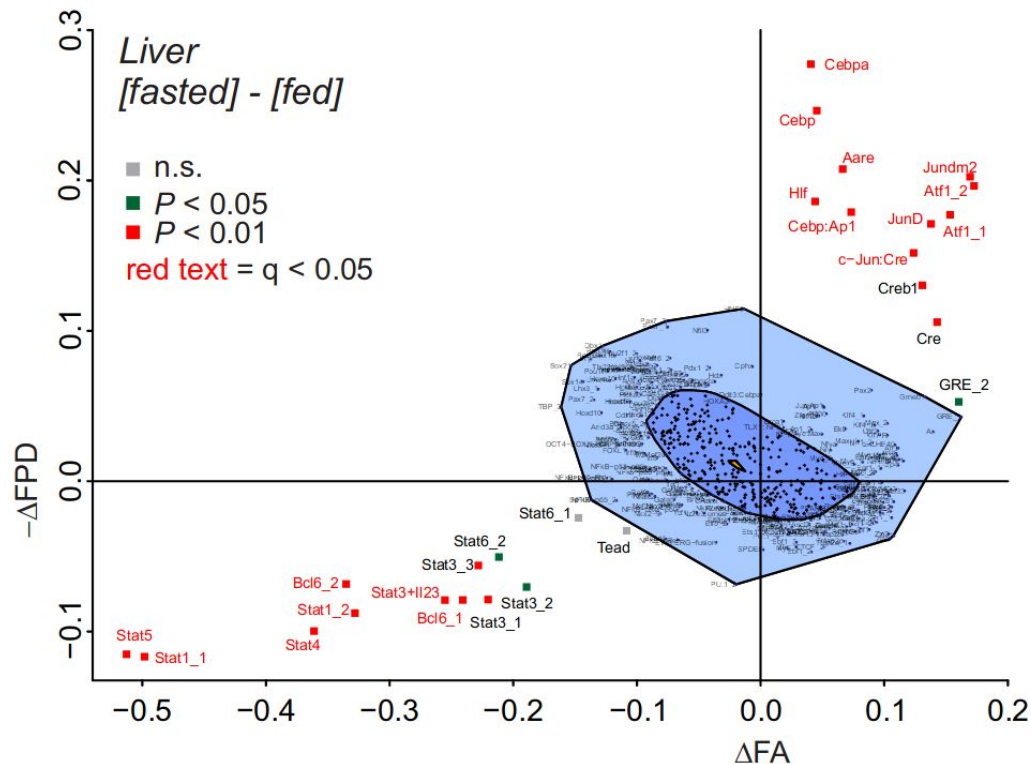
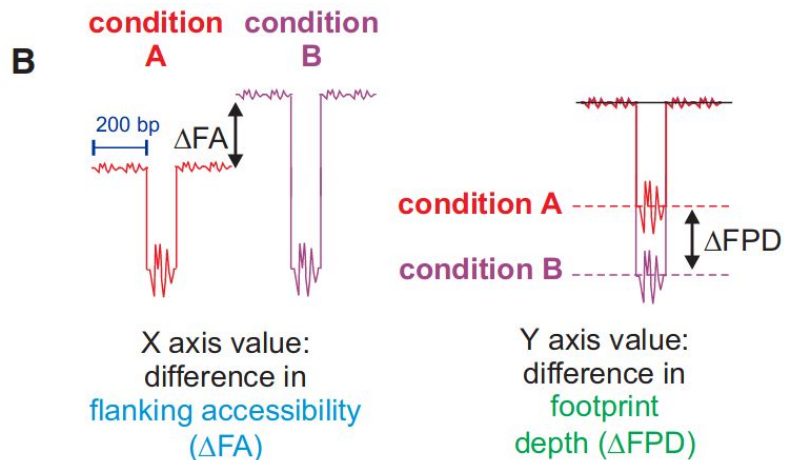
(At the moment, it's unclear which is best.)

Estimating TF activity from accessibility and footprints



(Baek, Goldstein and Hager, Cell Reports 2017)

Estimating TF activity from accessibility and footprints

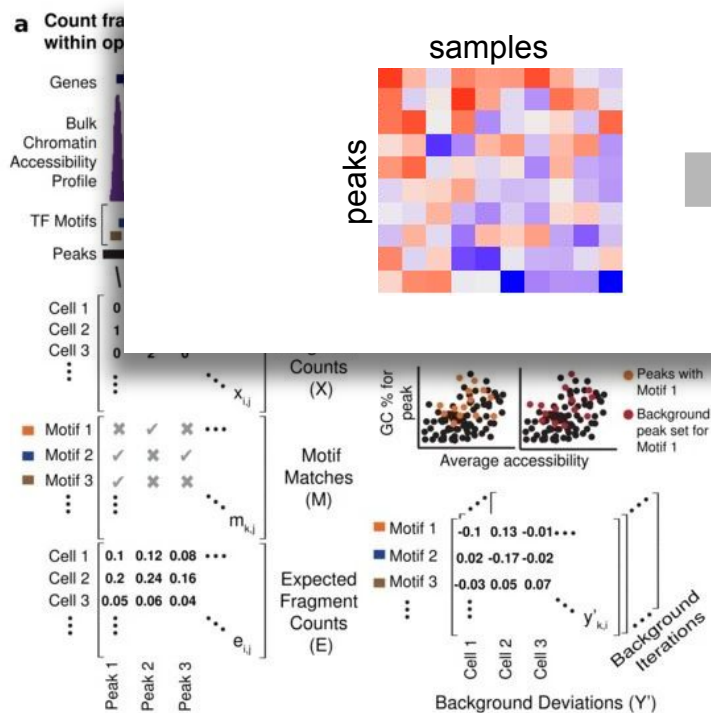


(Baek, Goldstein and Hager, Cell Reports 2017)

ChromVAR

[documentation](#)

chromVAR uses a simpler (and considerably faster) method, which essentially sums the counts of peaks that contain each motif, and compares this to a null distribution of similar peaks that don't.



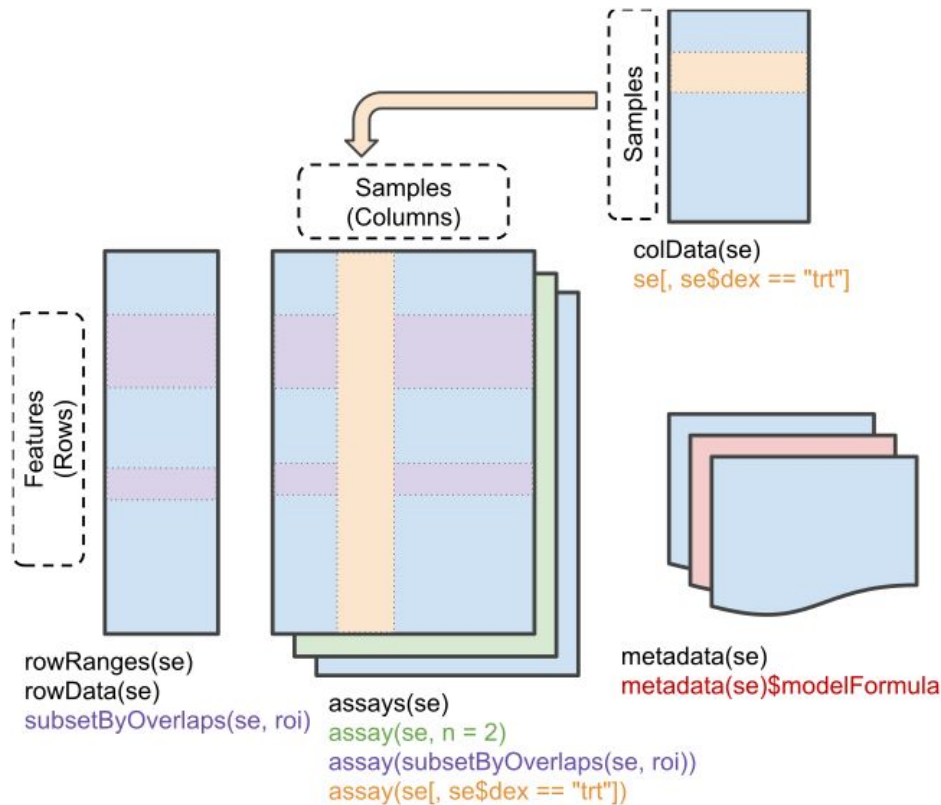
(adapted from Schep et al.,
Nature Methods 2017)

Although it's been developed especially for single-cell data, it's also routinely used for bulk.

Practical:

Motif accessibility analysis with chromVAR

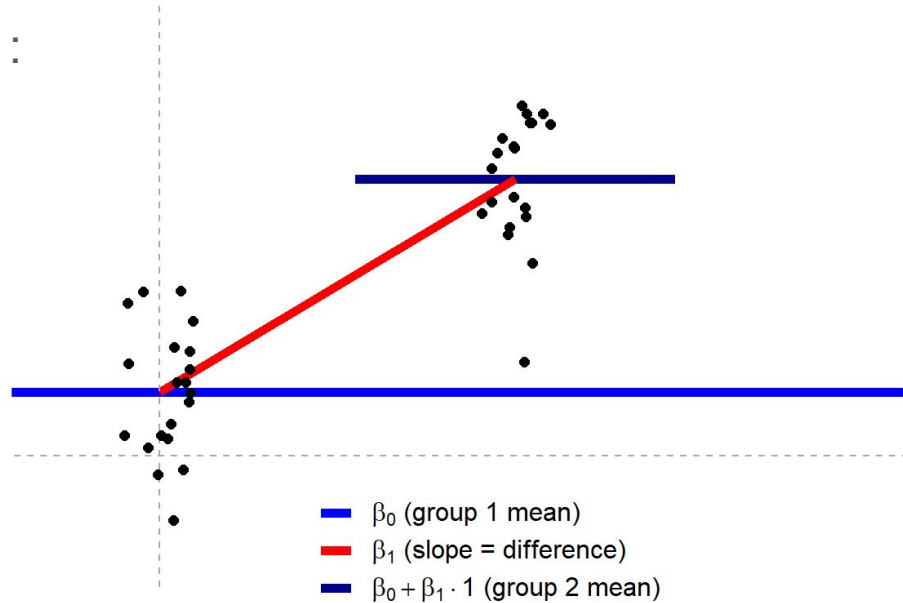
The SummarizedExperiment structure



Working with linear models

Most of the common statistical tests can be formulated as linear regression problems

Consider the t -test :



$$y \sim \beta_0 + x \cdot \beta_1$$

Where $x=0$ for group 1
and $x=1$ for group 2

(Taken from [an excellent explanation by Jonas Kristoffer Lindeløv](#))

Working with linear models - some simple examples

- For a two group comparison:
 - `~group` (equivalent to `~1+group`)
- Comparing between two groups, correcting for the effect of sex:
 - `~sex+group` → then we can still decide to test for the effects of the group by dropping that coefficient
- Finding sex-specific effects
 - `~sex*group` → equivalent to `~1+sex+group+sex:group`

Assignment

- Download (a subset of) ATAC-seq peak counts in the hippocampus upon stress (already in SummarizedExperiment format) :
 - https://ethz-ins.org/content/mouse_mm38_hippocampus.peakCounts.SE.rds
- Using this object, perform a chromVAR motif analysis, and run 2 differential motif accessibility analyses, respectively:
 - comparing stressed (denoted 'FSS' – forced swim stress) and control animals
 - comparing male and female animals
- For each analysis, report the top most significant motifs, plot a heatmap of the normalized accessibility scores across the samples for those motifs, and write a short paragraph interpreting the results.